**Supplementary Information**

**Supplementary Table 1:** Hyperparameters for model training

|  |  |
| --- | --- |
| Epoch | 12 |
| Batch Size | 2 |
| Learning Rate | 0.0025 |
| Loss function (LF)  LF for Classification  LF for Detection | Cross-entropy loss  L1 loss |
| Optimizer | Stochastic gradient descent |

**Supplementary Fig. 1** Representative annotated images. (A) Raw images; (B) annotation labels; (C) overlay images.

**Supplementary Fig. 2** Representative segmentation results for each test set. (A) Surgical instrument segmentation results, which are relatively good, even in images recorded using other laparoscopic systems. (B) Misrecognition of surgical forceps as T1 instead of T3. (C) Misrecognition of spleen as T2 in the LDG, and instability of the surgical tool recognition accuracy in LC and LPH (T1: harmonic shears; T2: endoscopic surgical electrocautery; T3: Aesculap AdTec atraumatic universal forceps; T4: Maryland; T5: Croce-olmi; T6: needle holder; LDG: laparoscopic distal gastrectomy; LC: laparoscopic cholecystectomy; LPH: laparoscopic partial hepatectomy).